# Report of the Groundfish Plan Team Retrospective Investigations Group

Part II: The compilation

September 2013 Dana Hanselman, Bill Clark, and Mike Sigler

#### Introduction

The retrospective investigations group (RIG) was formed at the request of the Groundfish Plan Team in November 2011. In September 2012, we (the RIG composed of Hanselman, Clark and Sigler) presented a document that summarized some past work on retrospective analyses, drawing largely from the report produced by the Woods Hole workshop on retrospective patterns (Legault 2009). The conclusions of the document were that there are no easy criteria to establish as to when a retrospective pattern is severe enough to render an assessment unacceptable. Several examples of Alaska stocks were shown and we concluded that retrospective patterns should be dealt with on a case-by-case basis and that patterns should be investigated by examining which parameters and data might be responsible for the pattern. We did not recommend that a retrospective pattern should necessarily be "fixed" as there is no way to be certain that the "fix" is any closer to the truth. Nevertheless, it is appropriate to investigate potential causes for assessments with substantial retrospective patterns. We suggested that a strong pattern could be used as a rationale to recommend a higher or lower ABC than otherwise would have been recommended. We asked assessment authors to run basic retrospective analyses for each stock as follows:

"For Alaska groundfish assessments with age-structured models (Tier 3 and higher), we recommend that a retrospective analysis be presented as part of the model evaluation. Specifically, stock assessment authors are requested to conduct the within-model approach and rerun the model, successively dropping data one year at a time. Specifically the analysis should include:

- 1. Running retrospectives back to 2002 (where 2002 would be a terminal year) for the base-case assessment in 2012 (i.e., drop 10 years of most recent data);
- 2. Plotting spawning biomass time series for each model run;
- 3. Plot of relative changes referenced to the terminal model run."

The majority of the assessment authors graciously complied with our request and provided us with results after the November 2012 Plan Team meeting. This document summarizes the retrospective patterns for 20 Alaska stock assessments and examines some features of specific assessments that may contribute to that pattern.

<sup>&</sup>lt;sup>1</sup> http://www.afsc.noaa.gov/refm/stocks/plan\_team/2012/sept/retrospective\_analysis.pdf

## **Methods**

Retrospective series of female spawning stock biomass were collected for 20 Alaska stocks from their most recent assessments (Table 1). We also compiled some descriptive statistics from each assessment to compare retrospective patterns to potential underlying causes (Table 1).

Several retrospective statistics were computed for each stock. One statistic originally proposed was by Mohn (1999) which has subsequently been revised (pers. comm. Robert Mohn via Grant Thompson). The original statistic was defined as the sum of relative differences between an estimated quantity from an assessment with a reduced time series and the same quantity estimated from the full time series. The difference between the (Mohn 1999) statistic and this revised version is that the average is used instead of the sum, so that the number of years of data left out ("peels") does not affect the magnitude. With a series of fits in which p = 1,..., P years of data are left out at the end, the revised statistic is:

$$\rho_{RMohn} = \sum_{p=1}^{P} \frac{X_{Y-p,p} - X_{Y-p,0}}{X_{Y-p,0}} / P \tag{1}$$

where Y is the last year in the full data series, p is the number of years at the end of the data series left out of a reduced series, and  $X_{y,p}$  denotes the estimate of some quantity (e.g., F or SSB) in year y obtained by fitting the model with the last p years of data left out ("peeled"). The estimate  $X_{y,p}$  is obtained with the full data series and is called the reference estimate. Mohn's statistic is the mean of relative differences between the terminal year estimates in the reduced fits and the corresponding estimates from the reference estimates (Equation 1). It will be zero when the peeled assessments match exactly with the reference assessment, or when the differences between the peeled assessments and the reference assessment are exactly balanced. The former case has no change from year to year, while the latter case would be characterized as exhibiting noise but not a retrospective bias. The Mohn p will become large, either positive or negative, when there is a consistent pattern of change in the peeled assessments relative to the reference assessment. There have not been rules of thumb developed regarding how large in absolute value Mohn p must be before an assessment is declared to exhibit an important retrospective bias.

Legault (2009) suggested a statistic (Wood's Hole  $\rho$ ) in which the deviations across all years in the peeled and reference fits are examined. This version examines whether historical estimates are also affected rather than only recent advice. We use the same general concept, but using averages rather than sums like equation (1):

$$\rho_{WH} = \left(\sum_{p=1}^{P} \left(\sum_{y=1}^{Y-p} \frac{X_{y,p} - X_{y,0}}{X_{y,0}}\right) / (Y-p)\right) / P \tag{2}$$

Parma (1993) used the root-mean-squared error of retrospective deviations because if errors were multiplicative and lognormal, the RMSE would be similar to the CV of the error. In that paper, only the terminal year RMSE was computed. For this study we compute RMSE for all years relative to the reference estimates. This measure is similar to the absolute value of the Wood's Hole  $\rho$  described above and measures overall retrospective variance:

$$RMSE = \sqrt{\left(\sum_{p=1}^{P} \left(\sum_{y=1}^{Y-p} (\log X_{y,p} - \log X_{y,0})^{2}\right) / (Y-p)\right) / P}$$
 (3)

Equation 1 measures only retrospective bias as far back as the model is peeled. An additional statistic of interest is the ratio of the year to year bias as measured by Mohn's revised  $\rho$  to the Wood's Hole  $\rho$ . This should give some measure of whether the retrospective bias is large across the whole model time series (<1), only recent estimates (>1), or if the bias

goes in the opposite direction in different parts of the model time series (<0). Note, that when both values of  $\rho$  are close to zero (no retrospective bias), the statistic is not very useful. We will call this Hanselman's  $\phi$ :

$$\phi_H = \frac{\rho_{RMohn}}{\rho_{WH}} \tag{4}$$

We hypothesized that certain features of the assessment or the input data compiled in Table 1 could contribute to the direction or magnitude of retrospective patterns. After the above statistics were calculated, we did a correlation analysis (Pearson's) to determine if any particular characteristics of a stock assessment might help explain why retrospective patterns occur.

#### **Results and Discussion**

Retrospective patterns relative to the reference model are plotted for all twenty stocks (Figure 1). Retrospective patterns showed both bias (i.e., consistent change in one direction) and variability (i.e., retrospective changes in estimates with no consistent direction). There was no consistent pattern of retrospective trends across stocks, but some properties emerged that included several stocks. For most rockfish stocks, the retrospective change occurred for the entire time series. This likely was caused by adding new imprecise survey data rather than model misspecification. Other stocks (AI\_POLL, GOA\_ATF, and BSAI\_NRS) had noisy patterns that appeared more related to model stability. Several stocks (BSAI\_POLL, GOA\_POLL, BSAI\_APL, and BSAI\_FHS) had very little retrospective bias or variability. Sablefish had very little absolute retrospective bias and variability, but had a very consistent pattern (Figure 1). Stocks were ranked with an overall retrospective measure that combines the average rank of the three measures of retrospective pattern (Table 2). Each statistical measure captures different patterns that are apparent when the stocks are plotted. Mohn's revised  $\rho$  is a measure that focuses only on the peeled estimates relative to the terminal estimate of SSB. This measure is likely the most relevant concerning current advice to management. The measure ranged from about -0.4 to 0.4 (Figure 2).

Species with the highest values of Mohn's revised  $\rho$  are compared and can be categorized into two types (Table 3). Three rockfish stocks have high absolute Mohn's revised  $\rho$  and can be called Forgone Yield Stocks (FYS). This means that in each successive year, the overall stock appears to be higher than previously estimated, meaning prior ABCs could have potentially been higher. Five other stocks can be categorized as Occasional Overage Potential Stocks (OOPS). If the recommended ABCs of these stocks were fully taken, the realized fishing mortality could have actually been higher than the maximum permissible fishing mortality. Several interesting patterns emerge when comparing modeling choices with these patterns. We examined asymptotic survey selectivity because it is often considered to be a conservative choice. We also examined what software was used for the stock assessment model. Models in Stock Synthesis are generally more likely to estimate non-asymptotic selectivities and a higher number of total parameters relative to AMAK-based models. None of the FYS had dome shaped selectivity for the survey, while 4/5 of the OOPS had dome-shaped survey selectivity (Table 3). In addition, 3/5 of the OOPS were using Stock Synthesis software, while all the FYS were using a version of AMAK software. The models in the OOPS categories on average estimated more effective parameters (total parameters minus fishing mortality and recruitment deviations) than the FYS models (93 vs. 22). The results suggest that avoiding dome-shaped survey selectivity and highly parameterized models might prevent the more perilous (OOPS) retrospective patterns.

Wood's Hole  $\rho$  and RMSE (Table 2, Figures 3 and 4) quantify retrospective patterns in a different way, by comparing all deviations from reference year estimates of SSB. In most cases, if Mohn's revised  $\rho$  was high (Table 2), so was Wood's Hole  $\rho$  as noted by Legault (2009). We used Hanselman's  $\phi$  as a way to show when they departed and when retrospective bias was recent or an overall change occurred in the series (Figure 5). Notable exceptions were GOA\_PCOD and SABLE, where Hanselman's  $\phi$  was high because almost all of the retrospective bias was only in the peeled years.

Conversely, Hanselman's  $\phi$  was below 1 for cases such as BSAI\_POLL and BSAI\_APL where there was very little overall retrospective variability. RMSE measures overall variability of the retrospective analysis, and identified stocks that may have models that are less constrained, or that have imprecise input data. An example of a relatively high RMSE stock, with relatively low retrospective bias is GOA\_POP. The large error deviations are likely caused by changes in imprecise data, but the time series pattern has both positive and negative retrospective deviations (Figure 1).

We conducted a correlation analysis (Table 4, Figure 6) of our measures of retrospective statistics (Table 2) with the various stock assessment features (Table 1) from all 20 assessments. The two measures of  $\rho$  are highly correlated with each other as expected (Figure 6). Significant correlations (p<0.05, uncorrected for multiple comparisons) were found between mean survey CV and all three measures of retrospective patterns, and natural mortality was positively correlated with Mohn's revised  $\rho$ . The higher the average survey CV the more likely the retrospective bias is negative, like the FYS described above, while the lower the survey CV the more likely it is to be an OOPS. A higher natural mortality rate is associated with a positive Mohn's revised  $\rho$  (an OOPS). Average fishing mortality had a similar and nearly significant association as with natural mortality. Note that the significant relationships with mean survey CV were not particularly linear (Figure 6), and could likely be better described with a curvilinear model. The loess smoothes in Figure 6 suggests that there is minimal retrospective bias at low survey CVs, positive bias at middle survey CVs, and negative bias at the highest survey CVs.

Underlying causes of these retrospective patterns are difficult to determine with certainty. Qualitatively, these rules of thumbs might be reasonable starting points with an example of a stock that has the characteristics:

- 1) High absolute Mohn's revised ρ: Misspecified natural mortality or selectivity, incorrect catch, high survey CV
  - Example: BSAI\_POP
- 2) High absolute Wood's Hole p: No contrast in catch history, paucity of historical data, low average-F
  - Example: GOA ATF
- 3) High RMSE: High survey CV, unconstrained parameters
  - Example: GOA NR
- 4) High Hanselman's φ: Strong historical data, high catch contrast.
  - Example: SABLE

We are not willing to recommend what constitutes an unacceptable level of retrospective pattern, but we believe this cross-stock comparison can suggest which assessments to examine in more detail. We recommend the assessments with the top 20% (4) ranked retrospective patterns are examined for sources of retrospective patterns with several suggestions based on our knowledge of the stock assessments. These include, but are not limited to:

- 1) GOA NR: Fix q and M at different values; Explore sensitivity to historical catch
- 2) BSAI POP: Fix q and M at different values; Examine effect of time-varying selectivity on retrospective patterns.
- 3) GOA\_ATF: Force both male and female selectivity to maximum of 1; Explore lower values of M; Explore estimating q
- 4) BSAI GT: Force asymptotic survey selectivities; Estimate M

# **Acknowledgments**

We wish to thank all of the stock assessment authors for providing retrospective run results to us, and Grant Thompson for providing access to the Species Information System files he compiles annually.

## References

- Legault, C.M., Chair. 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-01; 30 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at <a href="http://www.nefsc.noaa.gov/nefsc/publications/">http://www.nefsc.noaa.gov/nefsc/publications/</a>
- Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56: 473-488.
- Parma, A. N. 1993. Retrospective catch-at-age analysis of pacific halibut: implications on assessment of harvesting policies. In Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, pp. 247–265. Ed. by G. Kruse, D. M. Eggers, C. Pautzke, R. J. Marasco, and T. J. Quinn II. Alaska Sea Grant College Program.

Table 1. Alaska stocks used in retrospective analyses and some key features of each assessment that might affect retrospective trends. Mean survey CV is the average of coefficient of variations of individual survey biomass estimates (the length of the time series differs across areas and assessments). Survey time series CV is the coefficient of variation of all the point estimates of the survey time series. M is natural mortality. Recruit CV is the coefficient of variation of the recruitment time series from 1977-current. Average F is the average fully-selected fishing mortality from 1977-current.

		N.4	Survey		Danist	
Charle ID	Common nome	Mean	Time	N 4	Recruit.	۸۰۰۵۳۵۵۵ ۲
Stock ID	Common name	Survey CV	Series CV	М	CV	Average F
GOA_NR	GOA Northern rockfish	0.426	0.638	0.060	0.974	0.027
BSAI_POP	BSAI Pacific ocean perch	0.250	0.538	0.062	0.655	0.031
GOA_ATF	GOA Arrowtooth flounder	0.089	0.286	0.275	0.333	0.009
BSAI_GT	BSAI Greenland turbot	0.181	0.437	0.112	1.844	0.124
GOA_REBS	GOA Rougheye/blackspotted rockfish	0.176	0.146	0.030	0.668	0.024
GOA_DUSKY	GOA Dusky rockfish	0.397	0.538	0.070	0.924	0.053
BSAI_PCOD	BSAI Pacific cod	0.085	0.330	0.340	0.669	0.191
GOA_PCOD	GOA Pacific cod	0.177	0.364	0.340	0.430	0.228
BSAI_ATKA	BSAI Atka mackerel	0.249	0.375	0.300	0.609	0.167
GOA_POP	GOA Pacific ocean perch	0.250	0.432	0.060	0.680	0.102
BSAI_YFS	BSAI Yellowfin sole	0.089	0.202	0.120	0.502	0.079
BSAI_NRS	BSAI Northern rock sole	0.090	0.481	0.150	0.553	0.022
BSAI_NR	BSAI Northern rockfish	0.302	0.327	0.041	0.648	0.086
AI_POLL	AI Walleye pollock	0.345	0.767	0.180	2.130	0.076
BSAI_POLL	EBS walleye pollock	0.110	0.342	0.300	0.629	0.211
BSAI_BSRE	BSAI Blackspotted/rougheye rockfish	0.278	0.898	0.033	1.419	0.047
SABLE	Alaska sablefish	0.131	0.090	0.100	1.031	0.080
GOA_POLL	GOA Walleye pollock	0.176	0.973	0.300	1.212	0.141
BSAI_FHS	BSAI Flathead sole	0.111	0.276	0.200	0.516	0.066
BSAI_APL	BSAI Alaska plaice	0.114	0.160	0.130	0.418	0.045

Table 2. Comparison of statistics to examine retrospective patterns. The table is sorted by the average rank of the 1<sup>st</sup> 3 columns (absolute value for the two  $\rho$  statistics). RMSE is root-mean-square-error and H $\phi$  is Hanselman's  $\phi$ , the ratio of Mohn's revised  $\rho$  to Wood's Hole  $\rho$ . Rank is from most to least retrospective pattern. BSAI\_GT and GOA\_REBS were tied, but because of its higher Mohn's revised  $\rho$ , BSAI\_GT won the tiebreaker. Green is negative, red is positive.

Stock	Mohn's revised ρ	Wood's Hole ρ	RMSE	Нφ	Rank
GOA_NR	-0.443	-0.357	0.521	1.241	1
BSAI_POP	-0.401	-0.26	0.352	1.541	2
GOA_ATF	0.383	0.343	0.312	1.117	3
BSAI_GT	0.418	0.19	0.220	2.196	4
GOA_REBS	0.342	0.253	0.239	1.351	5
GOA_DUSKY	-0.304	-0.221	0.283	1.373	6
BSAI_PCOD	0.335	0.175	0.189	1.912	7
GOA_PCOD	0.422	0.068	0.128	6.191	8
BSAI_ATKA	0.255	0.14	0.179	1.824	9
GOA_POP	0.115	0.127	0.165	0.908	10
BSAI_YFS	-0.131	-0.065	0.099	2.014	11
BSAI_NRS	-0.101	-0.072	0.080	1.398	12
BSAI_NR	-0.083	-0.053	0.095	1.553	13
AI_POLL	0.086	0.048	0.058	1.801	14
BSAI_POLL	-0.019	-0.043	0.061	0.450	15
BSAI_BSRE	0.028	0.036	0.058	0.778	16
SABLE	0.089	0.016	0.043	5.404	17
GOA_POLL	0.037	0.023	0.052	1.636	18
BSAI_FHS	0.007	0.042	0.057	0.157	19
BSAI_APL	0.018	-0.006	0.015	-3.093	20

Table 3. Selection of stocks with highest absolute Mohn's revised  $\rho$ . Forgone yield stock (FYS) category means there was potential for higher utilization. Occasional Overage Potential Stock (OOPS) means it is possible for a fishing mortality higher than maximum permissible if ABC was fully taken. Effective parameters is total parameters minus fishing mortality deviations (F\_devs) and recruitment deviations (Rec\_devs).

	Forgone Yield Stock (FYS)			Occasional Overage Potential Stock (OOPS)				
Stock	GOA_NR	BSAI_POP	GOA_DUSKY	GOA_PCOD	BSAI_GT	GOA_ATF	GOA_REBS	BSAI_PCOD
Mohn's rev. $\rho$	-0.44	-0.40	-0.30	0.42	0.42	0.38	0.34	0.33
Wood's Hole $\rho$	-0.36	-0.26	-0.22	0.07	0.19	0.34	0.25	0.18
RMSE	0.52	0.35	0.28	0.13	0.22	0.31	0.24	0.19
Ηφ	1.24	1.54	1.37	6.19	2.20	1.12	1.35	1.91
Rank	1	2	6	8	4	3	5	7
Survey Selex	Asymp.	Asymp.	Asymp.	Dome	Dome	Asymp.	Dome	Dome
Software	AMAK	AMAK	AMAK	SS	SS	AMAK	AMAK	SS
Parameters	145	143	99	239	129	138	139	184
F_devs	51	53	35	-	-	51	35	-
Rec_devs	81	50	51	48	72	63	56	38
Eff. Param.	13	40	13	191	57	24	48	146

Table 4. Pearson's correlations (r) that exceed 0.20 between stock characteristics and retrospective statistics for all 20 Alaska stocks. Correlations that would be considered significant are in **bold** (p<0.05). Variables are defined in Table 1.

Stock statistic	Retrospective statistic	r	p-value
Mean.Survey.CV	Wood's hole $\rho$	-0.512	0.021
M	Mohn's revised $\rho$	0.486	0.030
Mean.Survey.CV	Mohn's revised $\rho$	-0.473	0.035
Mean.Survey.CV	RMSE	0.464	0.039
Average.F	Mohn's revised $\rho$	0.434	0.056
Average.F	Hanselman's φ	0.392	0.087
M	Wood's hole $\rho$	0.374	0.104
Average.F	RMSE	-0.300	0.198
Survey.Time.Series.CV	Wood's hole $\rho$	-0.284	0.225
Survey.Time.Series.CV	Mohn's revised $\rho$	-0.283	0.227
Average.F	Wood's hole $\rho$	0.215	0.363
M	Hanselman's φ	0.206	0.383

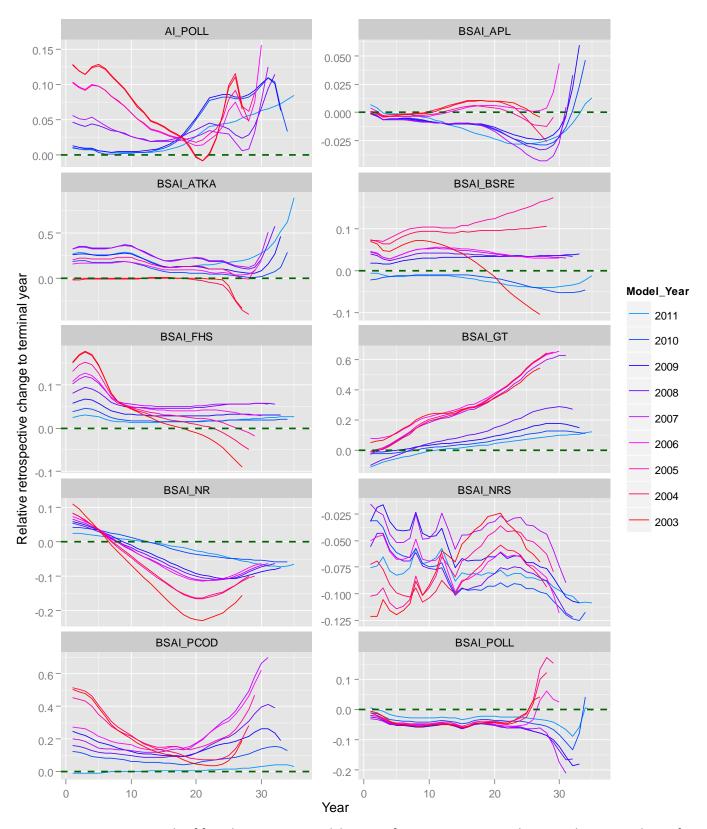


Figure 1a. Retrospective peels of female spawning stock biomass for nine past years relative to the terminal year for ten Alaska stocks. Note that the scales are different across stocks to examine patterns. Stock IDs are defined in Table 1. Years are from 1977-2012 for most stocks, GOA rockfish are from 1976-2011 because of the biennial assessment schedule. GOA\_REBS and GOA\_DUSKY are fixed where the SSB in 1976 is the same as 1977 because they do not estimate prior to 1977.

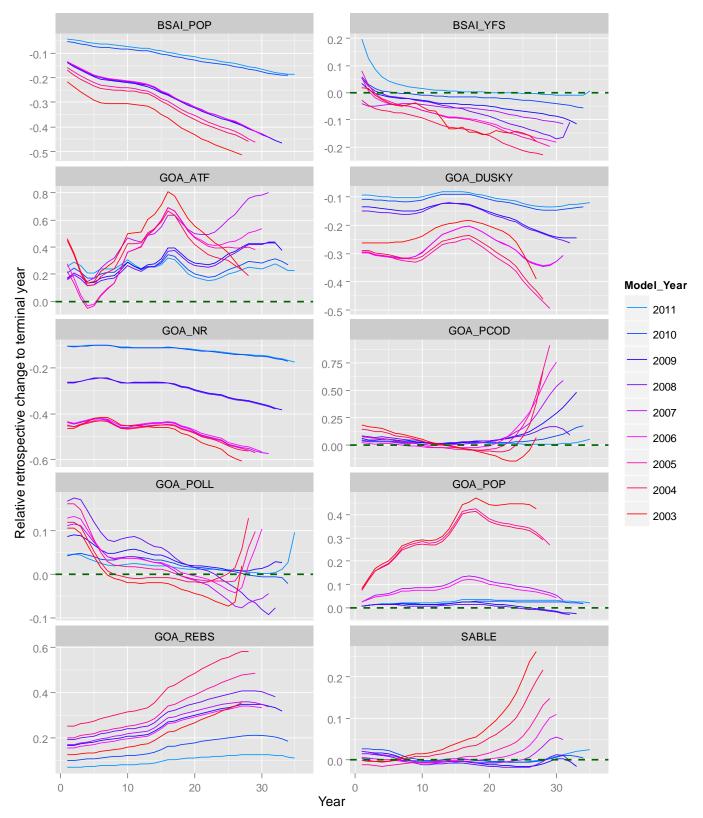


Figure 1b. Retrospective peels of female spawning stock biomass for nine past years relative to the terminal year for ten Alaska stocks. Note that the scales are different across stocks to examine patterns. Stock IDs are defined in Table 1. Years are from 1977-2012 for most stocks, GOA rockfish are from 1976-2011 because of the biennial assessment schedule. GOA\_REBS and GOA\_DUSKY are fixed where the SSB in 1976 is the same as 1977 because they do not estimate prior to 1977.

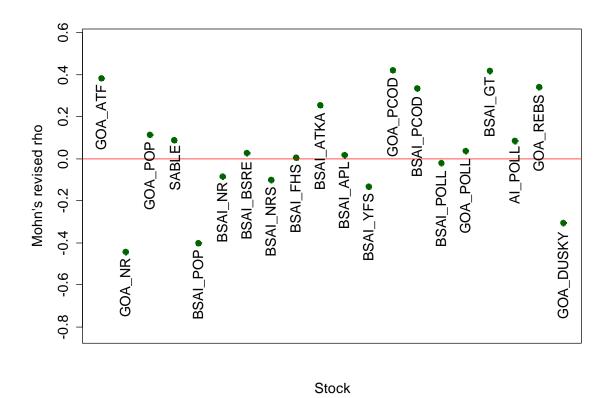


Figure 2. Mohn's revised  $\rho$  retrospective statistic on female spawning biomass compared across twenty Alaska stocks. Statistic is defined in text.

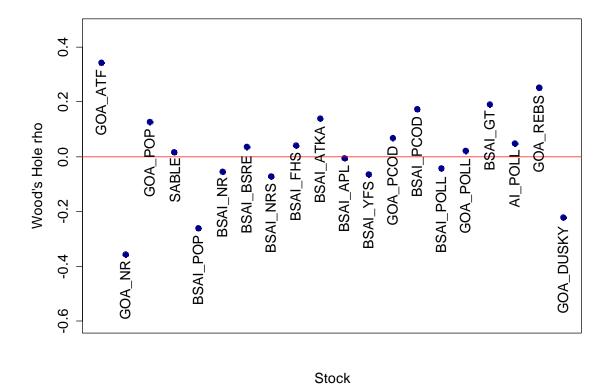


Figure 3. Wood's Hole  $\rho$  retrospective statistic on female spawning biomass compared across twenty Alaska stocks. Statistic is defined in text.

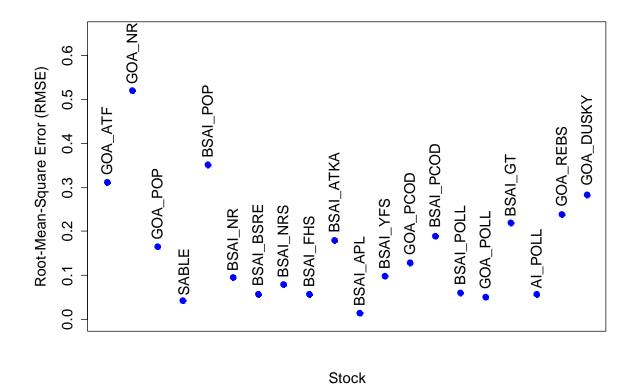


Figure 4. Root-mean-square-error (RMSE) retrospective statistic on female spawning biomass compared across twenty Alaska stocks. Statistic is defined in text.

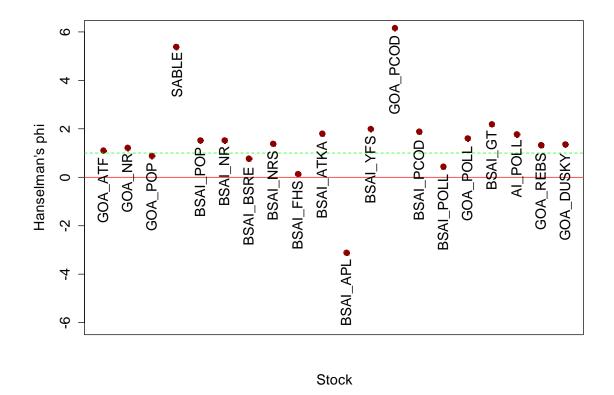


Figure 5. Hanselman's  $\phi$  retrospective statistic on female spawning biomass compared across twenty Alaska stocks. Hanselman's  $\phi$  is the ratio of Mohn's revised  $\rho$  to Wood's Hole  $\rho$ . The green dashed horizontal line is set at one when recent retrospective patterns are more severe than overall model retrospective patterns. The solid red line is set at zero, when  $\phi$  is below zero, the recent retrospective pattern is in the opposite direction of the overall pattern.

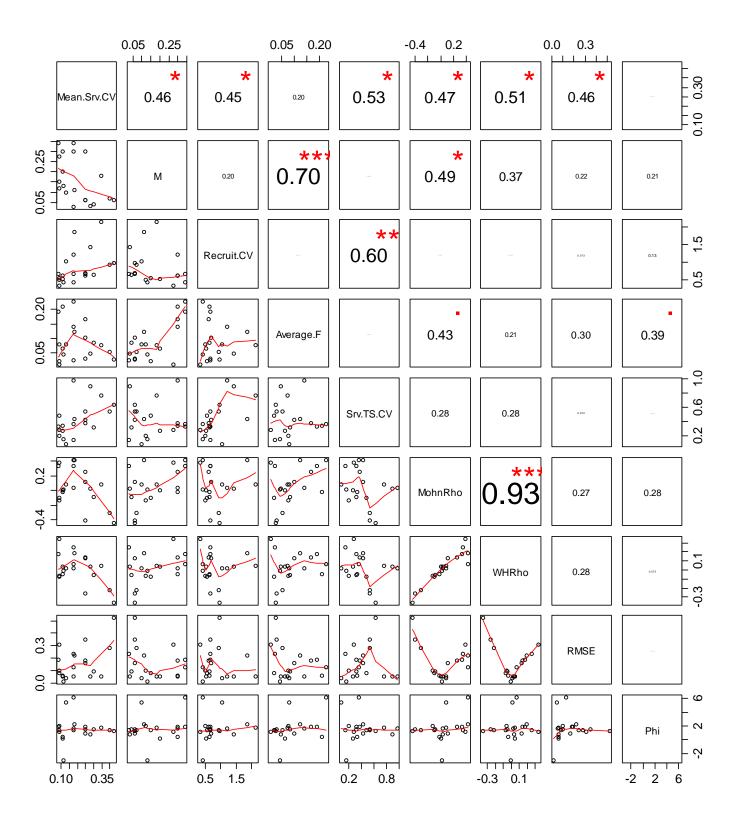


Figure 6. Correlation matrix of stock assessment statistics and retrospective statistics for twenty Alaska stocks. Red lines are loess smooths. Stars indicate significance (1-star=p<0.05, 2-star=p<0.01, 3-star=p<0.001). Variables are defined in Tables 1 and 2.